Chapter 31
Noise Characteristics of On-Chip Power Networks with Decoupling Capacitors

The high frequency response of a power distribution system is the focus of this chapter. The impedance of the power distribution system at high frequencies is determined by the characteristics of the on-chip power distribution network. The impedance of a power system at a specific on-chip location is determined by the local resistive, inductive, and capacitive characteristics of the on-chip network. In this chapter, the impedance characteristics of both the on-chip power interconnect and the decoupling capacitors are combined to evaluate the noise characteristics of a power network. The inductance of an on-chip power distribution network is shown under specific conditions to be a significant design issue in high speed integrated circuits.

As discussed in Chap. 7, the inductance of the on-chip power and ground interconnect affects the impedance characteristics at relatively high frequencies; specifically, from the chip-package resonance to the highest frequencies of interest. The on-chip interconnect is a part of the current loop from the on-chip decoupling capacitors to the package decoupling capacitors. Typically, the inductance of this current loop is dominated by other parts of the loop—the bonding solder bumps, package conductors, and package decoupling capacitors. This situation is changing with technology scaling, as discussed in Sect. 31.1. The propagation of the power supply noise through the on-chip power distribution network is discussed in Sect. 31.2. The on-chip interconnect also provides a current path between the on-chip decoupling capacitors and the load. As the switching speed of the load increases, the inductance of the on-chip power lines can degrade the effectiveness of the on-chip capacitors, as discussed in Sect. 31.3. The chapter concludes with a summary.

© Springer International Publishing Switzerland 2016

I.P. Vaisband et al., *On-Chip Power Delivery and Management*,
DOI 10.1007/978-3-319-29395-0_31

31.1 Scaling Effects in Chip-Package Resonance

The continuous improvement in the performance characteristics of integrated circuits is primarily due to decreasing feature sizes, as discussed in Chap. 1. Technology scaling, however, has highly unfavorable implications for the impedance characteristics of a power distribution system. The manner in which these scaling trends affect the impedance characteristics of a power distribution system are described in this section. Specifically, the impedance characteristics near the chip-package resonance is the topic of primary concern.

Ideal scaling theory is briefly reviewed in Sect. 5.1. The current density increases as S and the supply voltage decreases as $1/S$ in the ideal scaling scenario, as discussed in Chap. 5. To maintain the power noise margin at the same fraction of the power supply voltage, the impedance of the power distribution system will decrease as

$$Z_{\text{pds}} \propto \frac{V}{I} = \frac{1}{S^2 S_C^2}, \tag{31.1}$$

assuming that the circuit area increases by a factor S_C.

Power supply scaling is impeded in sub-100 nm technologies by the difficulties in reducing the transistor threshold voltages. A decrease in the power supply voltage therefore significantly deviates from the ideal scaling scenario [124]. Consider a scenario where the voltage levels are scaled by a factor S_V ($S_V < S$). The power supply V_{dd} decreases as $1/S_V$, while the transistor current I_{tr} scales as S/S_V^2. The current per circuit area I_{a} increases by a factor of $S^2 \cdot I_{\text{tr}} = S^3/S_V^2$. The impedance of the power supply system therefore decreases as

$$Z_{\text{pds}} \propto \frac{V}{I} \propto \frac{1}{S_V} \frac{1}{S_C^2 S^3/S_V^2} \propto \frac{S_V}{S} \frac{1}{S^2 S_C^2}. \tag{31.2}$$

This rate of decrease in the impedance is greater by a factor of S/S_V as compared to the ideal scaling scenario represented by (31.1).

The evolution of the impedance of a power distribution system in microprocessors is illustrated in Fig. 31.1. The rate of decrease in the impedance is approximately 2.7 times per technology generation (there are approximately four technology generations per decade). This rate is significantly greater than the dimension scaling factor $\sqrt{2}$, in good agreement with the scaling analysis characterized by (31.1) and (31.2). As described in Chap. 1, the rate of decrease in the target impedance has recently saturated (1.25 times per computer generation [286]). This decrease is due to the limited power dissipation capabilities of traditional air cooled packaging.

Consider the magnitude of the impedance at the frequency of the chip-package resonance, where the impedance is typically the greatest. The minimum impedance at the resonant frequency is the characteristic impedance of the tank circuit, $Z_0 = \sqrt{\frac{L}{C}}$, where C is the on-chip decoupling capacitance and L is the inductance

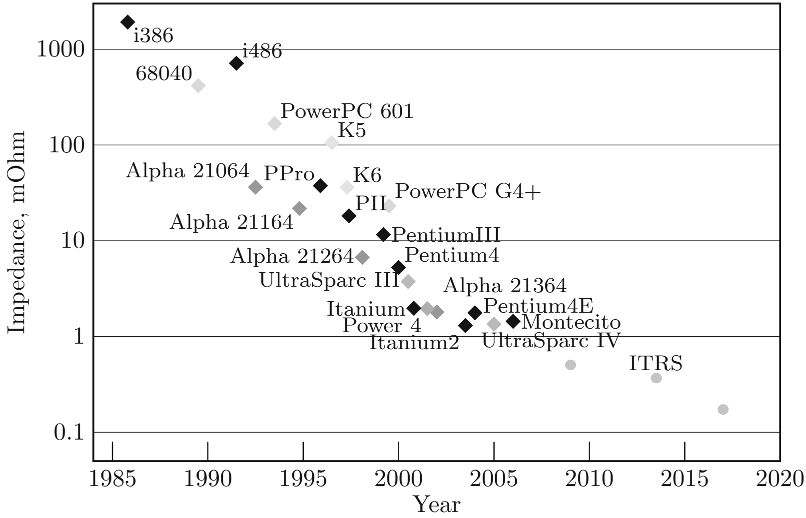

Fig. 31.1 Evolution of the impedance of a power distribution system in microprocessors. Several families of microprocessors and ITRS predictions [124] are shown in different shades of *gray*. The power supply noise margin is assumed to be 10 % of the power supply voltage

of the current loop from the on-chip load to the package decoupling capacitance, as discussed in Sect. 7.6. The decoupling capacitance per circuit area increases by a factor of S, and the overall capacitance of a circuit increases as $S_C^2 S$. The flip-chip contact density increases by a factor of S, assuming a contact pitch scaling factor of $\sqrt{S}$, as discussed in Chap. 5. Assuming a proportional decrease in the other components of the resonant inductance L, such as the inductance of the package conductors and the series inductance of the package capacitors, the overall loop inductance L decreases by a factor of S. Including an increase in the chip area as a factor of S_C^2, the inductance decreases by a factor of $1/S_C^2 S$. In this scenario, the resonant impedance only decreases by a factor of $\sqrt{\frac{1/S_C^2 S}{S_C^2 S}} = 1/S_C^2 S$, which is smaller than the requirements determined by (31.1) and (31.2). This reduction in impedance is therefore insufficient to satisfy a target noise margin. Further improvements in the circuit characteristics are necessary to approach the target specifications. The on-chip decoupling capacitance C is expensive to increase. When the available on-chip area is filled with decoupling capacitors, any additional decoupling capacitors increase die area and, consequently, the overall cost. Furthermore, in sub-100 nm technologies, the on-chip decoupling capacitors increase the static power consumption due to gate tunneling leakage current [517, 518].

The inductance of the current loop between the on-chip circuits and the package capacitors should be decreased to achieve the target impedance. The required decrease in inductance is particularly significant due to the square root dependence of the impedance on inductance. The inductance is reduced in advanced packaging

technologies through improvements in the structure of the package and package decoupling capacitors. Application of finely spaced metal layers and replacing the solder bump connections with denser microvia contacts have been described to achieve this objective [143].

Due to the aggressive reduction in the package inductance, the inductance of the off-chip portion of the current loop becomes comparable to the inductance of the on-chip power interconnect structures. This trend is in agreement with the general approach of using a system of hierarchical decoupling capacitors to achieve a low impedance power distribution system. The upper frequency limit of the low impedance characteristics of a power distribution system should be extended as the switching speed of the on-chip circuitry increases with technology scaling. Exclusively using on-chip capacitors to improve the high frequency impedance characteristics is a relatively expensive solution. Reducing the package inductance typically offers a more economical alternative by extending the frequency range of the package decoupling capacitors, thereby relaxing the requirements placed on the on-chip decoupling capacitors.

31.2 Propagation of Power Distribution Noise

The one-dimensional model described in Chap. 7 is inadequate to accurately describe the high frequency operation of a distributed circuit. As discussed in Chap. 8, the high frequency behavior of the on-chip power distribution network cannot be adequately described by a lumped model. A disturbance in the power supply voltage due to switching a local load propagates relatively slowly through the on-chip power distribution network. The power supply voltage is consequently non-uniform across the circuit die. These important effects are absent in a one-dimensional model. A two-dimensional model of the power distribution network is essential to accurately capture the high frequency impedance characteristics. The propagation of the power distribution noise through the on-chip power distribution network based on a simplified circuit model is discussed in this section.

The speed of noise propagation is an important characteristic of a power distribution network. The propagation speed of an undamped signal can be estimated using an idealized model of an on-chip power distribution network, where the decoupling capacitance is assumed to be uniformly distributed across a uniform power distribution grid. Assuming one-dimensional signal propagation, the power distribution grid is analogous to a capacitively loaded transmission line. The corresponding velocity of the signal propagation is

$$v = \frac{1}{\sqrt{L_{\square} C_{\mathrm{a}}}}, \tag{31.3}$$

where $L_{\square}$ is the sheet inductance of the power distribution grid (as described in Chap. 28) and C_{a} is the area density of the on-chip decoupling capacitance.

As a practical example, assume a global power distribution grid consists of two layers with mutually perpendicular lines. The width and pitch of the grid lines are 50 and 100 μm, respectively, similar to the characteristics of the upper layer of the two layer grid considered in Sect. 35.2. The corresponding sheet inductance is approximately $0.2\,\frac{\text{nH}}{\Box}$. A typical decoupling capacitance density C_{a} in high performance digital circuits manufactured in a 130 nm CMOS process is approximately $2\,\text{nF}^2/\text{mm}$. The velocity of the signal propagation based on these characteristics is approximately 1.6 mm/ns. This velocity is two orders of magnitude smaller than the speed of light in the circuit dielectric—approximately 150 mm/ns in silicon dioxide. The low velocity of the signal propagation is due to the high capacitance across the power and ground interconnect. This estimate is the upper bound on the signal velocity, as the resistance of the grid is neglected in (31.3). The resistance of the power lines further reduces the velocity of the signal. In overdamped power distribution networks, the signal propagation is determined by an *RC* rather than an *LC* time constant and approaches a diffusive *RC*-like signal behavior.

The relatively slow propagation of the power distribution noise has important circuit implications. From the perspective of a switching circuit, the low velocity noise propagation means that only the decoupling capacitance in the immediate proximity of the switching load is effective in limiting power supply variations at the load terminals. No charge sharing occurs during the switching transient between the load and the decoupling capacitors located farther than the propagation velocity times the switching time of the load, as described in Chap. 12. Alternatively, from the perspective of a quiescent circuit, the low propagation velocity means that the power supply level of the circuit is only affected by the switching loads that are located in close proximity to the circuit.

The idealized uniform model can also be used to estimate the inductive behavior of the on-chip power distribution grid. For one-dimensional signal propagation, the metric of inductive behavior for transmission lines described by (2.40) can be applied, yielding

$$\frac{t_r}{2\sqrt{L_{\Box}C_{\text{a}}}} < l < \frac{2}{R_{\Box}}\sqrt{\frac{L_{\Box}}{C_{\text{a}}}}, \tag{31.4}$$

where $R_{\Box}$ is the sheet resistance of the grid. Assuming a sheet resistance of $0.2\,\Omega/\Box$ and a signal rise time t_r of the load current of 100 ps,

$$0.1\,\text{mm} < l < 3\,\text{mm}, \tag{31.5}$$

the power supply noise exhibits a significant inductive component only within a limited distance from the switching load. In other terms, the damping factor ζ of the current path within a power distribution grid,

$$\zeta = \frac{R_{\Box}l}{2}\sqrt{\frac{C_{\text{a}}}{L_{\Box}}}, \tag{31.6}$$

is smaller than unity if the path length is smaller than 3 mm. Within this distance from the load, the response of a power distribution network is underdamped and the power supply noise can exhibit significant ringing. At greater distances from the load, the propagation of the power supply noise approaches a diffusive *RC*-like behavior.

The inductance of a pair of wide global power and ground lines is comparable to that of an on-chip signal line. The capacitive load of the power lines, however, is approximately three orders of magnitude greater, while the resistance is ten to a hundred times lower. The range of length where the power interconnect exhibits inductive behavior, as indicated by (31.5), is similar to that of an on-chip signal line.

Note, however, that the characteristic impedance of a power-ground line pair is approximately two orders of magnitude lower than the characteristic impedance of an on-chip signal path. The magnitude of the power distribution noise induced by switching an on-chip signal line is therefore two orders of magnitude smaller than the swing of a signal line transition.

31.3 Local Inductive Behavior

The idealized model used in the preceding section provides a reasonable approximation of the noise propagation at a relatively large geometric scale, i.e., where the wavelength of the signal is significantly larger than the pitch of the power lines. At smaller scales (and, consequently, shorter propagation times), the discrete nature of both the power load and the decoupling capacitors may be significant under certain conditions. Particularly, the high frequency characteristics of the power distribution interconnect become crucial. These local effects are discussed in this section.

A low impedance power distribution system during and immediately after the switching of an on-chip load is maintained using on-chip decoupling capacitors. The on-chip decoupling capacitance limits the variation of the power supply until the package decoupling capacitors become effective. As discussed in Sect. 11.3, the intrinsic parasitic capacitance of the load circuit typically provides a small fraction of the required decoupling capacitance. The intrinsic capacitance is embedded in the circuit structure. Consequently, the impedance between the intrinsic capacitance and the switching load capacitance is small. The intentional decoupling capacitors augment the intrinsic capacitance of the circuit to reach the required level of capacitance. The intentional capacitance, however, is typically added at the final stages of the circuit design process and is often physically located at a significant distance from the switching load. As the switching time of the load decreases, the impedance of the power interconnect becomes increasingly important. The significance of the power line impedance on the efficacy of the decoupling capacitors is demonstrated in the following example.

Consider an integrated circuit manufactured in a sub-100 nm CMOS technology. A high power local circuit macro, $200 \times 200\,\mu\text{m}$ in size, switches a 20 pF load

capacitance C_{load} within a 100 ps time period t_r. Assuming a 1 V power supply, the maximum power current of the circuit is estimated as

$$I_{\max} \approx \frac{C_{load} V_{\mathrm{dd}}}{t_r/2} = \frac{20\,\mathrm{pF} \times 1\,\mathrm{V}}{100\,\mathrm{ps}/2} = 400\,\mathrm{mA}, \tag{31.7}$$

and the maximum current transient as

$$\left(\frac{dI}{dt}\right)_{\max} \approx \frac{I_{\max}}{t_r/2} = \frac{400\,\mathrm{mA}}{100\,\mathrm{ps}/2} = 8 \times 10^9\,\frac{\mathrm{A}}{\mathrm{s}}. \tag{31.8}$$

The decoupling capacitance embedded within the circuit is assumed to be insufficient, supplying only half of the required current. The rest of the current is supplied by the nearest on-chip decoupling capacitor. To limit the resistive and inductive voltage drops to below 100 mV, 10 % of the power supply in this 1 V system, the resistance and inductance of the current path between the load circuit and the decoupling capacitor should be smaller than, respectively,

$$R_{\max} = \frac{0.1 V_{\mathrm{dd}}}{0.5 I_{\max}} = \frac{0.1\,\mathrm{V}}{0.2\,\mathrm{A}} = 0.5\,\Omega \tag{31.9}$$

and

$$L_{\max} = \frac{0.1 V_{\mathrm{dd}}}{0.5\,(dI/dt)_{\max}} = \frac{0.1\,\mathrm{V}}{4 \times 10^9\,\mathrm{A/s}} = 25\,\mathrm{pH}. \tag{31.10}$$

These impedance specifications are demanding. Assume that the physical distance between the load and the capacitor is 100 µm. Consider a scenario where the load and capacitor are connected by two global power and ground lines that are 50 µm wide, 1 µm thick, and are placed on a 100 µm pitch, as illustrated in Fig. 31.2a. The resistance of the current path is approximately 0.08 Ω, well below the limit set by (31.9). The inductance of the path, however, is approximately 80 pH, exceeding the limit set by (31.10).

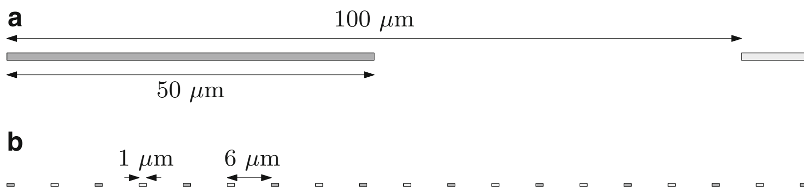

Fig. 31.2 Cross section of a current path connecting the load and decoupling capacitance. The power lines are shown in a *darker gray*, while the ground lines are shown in a *lighter gray*. The connection between the load and decoupling capacitance can be made using either (**a**) thick and wide global power lines or (**b**) finer local power lines. The dimensions are drawn to scale

Alternatively, if the load and decoupling capacitors are connected by fine lines of a local distribution network, for example, by 32 interdigitated power and ground lines with a $0.4 \times 1\,\mu m$ cross section and a $6\,\mu m$ line pitch, as illustrated in Fig. 31.2b, the inductance of the current path is reduced to 7 pH, well below the limit set by (31.10). The resistance in this case, however, is approximately $0.63\,\Omega$, exceeding the limit set by (31.9). The high current density in these fine lines is also likely to violate existing electromigration reliability constraints.

The target impedance characteristics are therefore more readily achieved if both wide and thick lines in the upper metal layers and fine lines in the lower metal layers are used. The superior impedance characteristics of such interconnect structures are described in Chap. 35. Alternatively, the width of the global power and ground lines in the upper metal layers should be greatly decreased. This approach will decrease the inductance of the grid, at the expense of a moderate increase in resistance, as discussed in Chap. 30.

Due to the limitations of the power interconnect, the on-chip decoupling capacitance should be placed in the immediate vicinity of the switching load in order to be effective. This requirement necessitates novel approaches to allocating the on-chip decoupling capacitance, as described in Chap. 12. A common design approach, where the bulk of the decoupling capacitance is placed among the circuit blocks after the initial design of the blocks has been completed, as shown in Fig. 11.18, does not permit placing the decoupling capacitors sufficiently close to the circuits far from the block boundary. As the feature size of the on-chip circuits decreases, the capacitance allocation process should be performed at a commensurately finer scale, as schematically illustrated in Fig. 31.3.

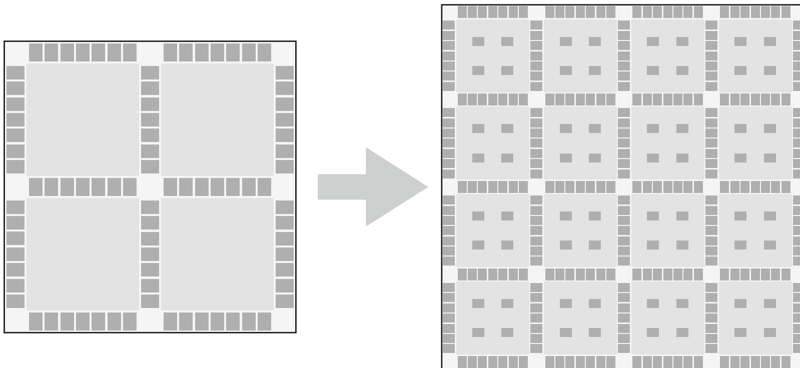

Fig. 31.3 The effect of circuit scaling on allocation of the on-chip decoupling capacitance. The circuit blocks are shown in *darker gray*, the decoupling capacitors are shown in *lighter gray*. As the circuit feature sizes decrease, the allocation of the on-chip decoupling capacitance should be performed at a commensurately finer scale. If the size of the circuit blocks does not decrease in proportion to the feature size, the decoupling capacitors are placed within the circuit blocks, as shown on the *right*

The high frequency characteristics of the on-chip power interconnect are particularly important when the power load is non-uniformly distributed. Those circuits with the greatest peak power consumption require a significant decoupling capacitance in close proximity while the surrounding lower power circuits tend to require a relatively low decoupling capacitance.

The power consumption is particularly non-uniform in high performance digital circuits with a highly irregular structure, such as microprocessors, which are comprised of both low and high power circuit blocks. More than half of the die area in state-of-the-art microprocessors is occupied by memory circuit structures, which are characterized by a low switching activity and, consequently, low power consumption. The power density of a microprocessor core can be an order of magnitude higher as compared to the memory arrays; the core and synchronization circuits dissipate the dominant share of the overall circuit power. The distribution of the power consumption within the high power blocks is also typically non-uniform. The peak power demand circuitry with high load capacitances and switching activities, such as the arithmetic units and bus drivers, can exceed severalfold the worst case requirements of the surrounding circuits.

Contemporary trends in circuit design exacerbate the uneven distribution of the dissipated power. Similar to microprocessors, system-on-chip circuits integrate diverse circuit structures and also tend to exhibit a highly non-uniform power distribution pattern. Since the power consumption of integrated circuits has become a primary design priority, as described in Chap. 1, aggressive power saving techniques have become mandatory. Clock and power gating have gained wider use in order to decrease dynamic and leakage power consumption, respectively, in idle circuit blocks [308, 519–531]. While the power consumed by the circuit blocks is greatly reduced in the gated mode, abrupt transients in the power current are induced when a circuit block transitions from a power saving mode to active operation or vice versa. Power gating presents particular challenges to the analysis and verification of power distribution networks. A significant share of the decoupling capacitance is often disconnected from the global network during a power-down mode. This change in the decoupling capacitance can potentially cause power integrity problems in the surrounding circuits.

31.4 Summary

The effect of the decoupling capacitance and the inductance of on-chip interconnect on the high frequency impedance characteristics of a power distribution system is discussed in this chapter. The primary conclusions are summarized as follows.

- The inductance of the on-chip interconnect becomes more significant as the inductance of the package conductors is reduced
- The power noise propagates through the on-chip power distribution network at a relatively low velocity

- The response of the on-chip power distribution network is underdamped in close proximity to the load
- The impedance of the current path between the on-chip load and the on-chip decoupling capacitors becomes a critical design parameter as the power supply and circuit switching times decrease
- Allocating the on-chip decoupling capacitance should be performed at a finer scale as the feature size of the on-chip circuits decreases